

The Lindeberg theorem for Gibbs-Markov dynamics

Manfred Denker, Samuel Senti, Xuan Zhang

October 17, 2016

Abstract

A dynamical array consists of a family of functions $\{f_{n,i} : 1 \leq i \leq k_n, n \geq 1\}$ and a family of initial times $\{\tau_{n,i} : 1 \leq i \leq k_n, n \geq 1\}$. For a dynamical system (X, T) we identify distributional limits for sums of the form

$$S_n = \frac{1}{s_n} \sum_{i=1}^{k_n} [f_{n,i} \circ T^{\tau_{n,i}} - a_{n,i}] \quad n \geq 1$$

for suitable (non-random) constants $s_n > 0$ and $a_{n,i} \in \mathbb{R}$. We derive a Lindeberg-type central limit theorem for dynamical arrays. Applications include new central limit theorems for functions which are not locally Lipschitz continuous and central limit theorems for statistical functions of time series obtained from Gibbs-Markov systems. Our results, which hold for more general dynamics, are stated in the context of Gibbs-Markov dynamical systems for convenience.

1 Introduction

Probabilistic methods have been used for a long time in connection with number theory and some of these applications have formulations in terms of dynamical systems, but it took more than 50 years to realize the general importance for dynamics. Continued fraction is a typical example of such a common approach in number theory and dynamics. While the ergodic theorem has a direct counterpart in Kolmogorov's strong law of large numbers, the classical de Moivre-Laplace central limit theorem describing its fluctuation about the mean has none. Today, central limit theorems (CLT) are widespread in the study of fluctuations of Birkhoff ergodic sums in dynamical systems. Ideas borrowed from probability theory such as stationary mixing processes [19] and Gordin's martingale-coboundary method [15] are commonly used to prove central limit theorems. The survey paper [8] has a comprehensive list of references up to 1986. More recently, Chazottes [4] reviewed probabilistic laws for ergodic sums in dynamical systems modeled by Young towers ([25]). A central limit theorem for Markov fibred systems with the Schweiger property was proven in [3]. Examples of such systems include parabolic rational maps, Young towers and

Gibbs-Markov maps. A CLT for general rational maps was proven in [11] using Gordin's method. All these results are concerned with Birkhoff sums. In order to obtain a CLT, the observables in the Birkhoff sums are usually assumed to be Hölder continuous. This is partly due to the popular spectral gap method, which usually holds on Banach spaces endowed with Hölder norms. On the other hand it is still an open problem to determine the class of functions in L^2 satisfying the CLT.

From an applied viewpoint, ergodic sums provide only a limited method to draw conclusions on a dynamical system. A much wider approach is formulated in terms of design of experiments where different time series and their interplay are considered. This leads to the need of analyzing arrays containing different ergodic sums. To our knowledge this concept was first formulated for maps of the interval and some special statistical functionals in [7]. Recently, [17] used a special form of such an approach to obtain CLT for shrinking targets. In other directions, one should also note CLTs in the settings of random dynamics or sequential dynamics such as in [6, 5, 21]. Lindeberg's central limit theorem deals with arrays of independent random variables, i.e. families of random variables defined on row-wise different probability spaces. We formulate Lindeberg's central limit theorem for dynamical arrays, and prove CLTs for arrays in dynamical systems, here Gibbs-Markov maps. Examples include certain countable state Markov chains and Markov maps of the unit interval given in [2] as well as parabolic rational maps in [3]. Other examples can be found in [1]. We use two classical methods: the characteristic operators approach as in [23] and Lindeberg's method as in [22] for blocks to prove CLTs. It will be clear from the proof that our results can be extended to more general systems since only spectral properties of transfer operators and metric properties are taken from Gibbs-Markov systems. It is also clear how to extend the results to Young towers over Gibbs-Markov maps. Recent development can be found in [24]. For simplicity, we keep our discussions restricted to Gibbs-Markov systems.

Dynamical arrays have many applications. For instance, one may use an array of Hölder continuous observables to approximate Birkhoff sums of observables of lesser regularity. An example is given in Corollary 4.2. In comparison, note that Gouëzel [16] proved a CLT for Birkhoff sums of observables with Hölder norm in L^η , where $0 < \eta < 1$. In another paper [12] we have used CLTs for arrays to study fluctuations for ergodic sums over periodic orbits. Another possible application is through coupling Birkhoff sums of different dynamical systems.

We recall some background material on Gibbs-Markov systems and spectral properties of their transfer operators in Sections 2 and 3. Section 4 contains a central limit theorem (Theorem 4.1) for sequences of Birkhoff sums $\sum_{i=1}^{k_n} f_n \circ T^i$ ($n \geq 1$). Although this is a special dynamical array the central limit theorem is treated separately since the method of proof is different from the other main theorem and may have future applications to other dynamical systems. Such theorems provide central limit theorems for Birkhoff sums $\sum_{i=1}^n f \circ T^i$ for certain functions which are not Lipschitz (or Hölder) continuous. We provide one easy example and others are not difficult to obtain. Section 5 contains the Lindeberg

central limit theorem (Theorem 5.3) for dynamical arrays. Here we deal with the central limit theorem in its most general form as loosely formulated in the abstract. The precise statement and assumptions are presented at the beginning of this section. There are many applications of this theorem. In Section 6 we provide one of them by showing the asymptotic normality of the Wilcoxon two sample rank statistics. Other examples are given in [26] and will be derived elsewhere. To get an idea of the scope of other possible applications, one may consult [14] or similar expositions.

2 Gibbs-Markov systems

Gibbs-Markov systems were first formulated in [2]. Let $(\Omega, \mathcal{B}, \mu, T)$ denote a nonsingular transformation of a standard probability space. Consider a countable partition α of Ω mod μ , $\alpha = \{a_i : i \in I\}$. For $x, y \in \Omega$, define

$$s(x, y) := \min_{n \in \mathbb{N}} \{n + 1 : T^n(x) \text{ and } T^n(y) \text{ belong to different elements of } \alpha\}.$$

For any $r \in (0, 1)$, set $r(x, y) := r^{s(x, y)}$ on Ω , which will become a metric. We use the same letter r to express the dependence of the metric on the choice of r . It will be clear in the subsequent context when does r represent a number or a metric.

Definition 2.1. A quintuple $(\Omega, \mathcal{B}, \mu, T, \alpha)$ is called a Gibbs-Markov map (or system) if the following four conditions hold modulo μ .

1. α is a strong generator of \mathcal{B} by T , i.e. $\sigma(\{T^{-n}\alpha : n \geq 0\}) = \mathcal{B}$.
2. For every $a \in \alpha$, $Ta \in \sigma(\alpha)$ and the restriction $T|_a$ is invertible and (two-sided) nonsingular.
3. $\inf_{a \in \alpha} \mu(Ta) > 0$.
4. For each $n \geq 1$ and $a \in \bigvee_{i=0}^{n-1} T^{-i}\alpha$, denote the μ -nonsingular inverse branch $T^{-n}|_{T^n a}$ by $v_a : T^n a \rightarrow a$ and its Radon-Nikodym derivative by v'_a . Then, there exist $r \in (0, 1)$ and $M > 0$ such that for any $n \geq 1, a \in \bigvee_{i=0}^{n-1} T^{-i}\alpha$ and $x, y \in T^n a$,

$$\left| \frac{v'_a(x)}{v'_a(y)} - 1 \right| \leq M \cdot r(x, y). \quad (1)$$

Remark 2.2. Usually we do not specify \mathcal{B} and write only (Ω, μ, T, α) . Also note that a number r and hence the metric $r(\cdot, \cdot)$ are determined within the definition of a Gibbs-Markov map.

We will work with the following Banach spaces: given a Gibbs-Markov system (Ω, μ, T, α) and any partition ρ of Ω , the Hölder norm of a function

$f : \Omega \rightarrow \mathbb{C}$ subject to the partition ρ is defined by

$$D_\rho f := \sup_{b \in \rho} \sup_{x, y \in b, x \neq y} \frac{|f(x) - f(y)|}{r(x, y)},$$

where sup is understood to be taken μ almost everywhere. Denote the usual L^q -norm by $\|\cdot\|_q$, $1 \leq q \leq \infty$. Then

$$\|f\|_{\infty, \rho} := \|f\|_\infty + D_\rho f$$

defines a larger norm. Denote the subspace consisting of functions of finite $\|\cdot\|_{\infty, \rho}$ norm by L_ρ^∞ . It is standard to show that L_ρ^∞ is a Banach space.

Remark 2.3. Throughout this paper we will always assume that (Ω, μ, T, α) is a topologically mixing and measure-preserving Gibbs-Markov system, namely for any $a, b \in \alpha$, there is $n_{a,b} \in \mathbb{N}$ such that for every $n > n_{a,b}$, $b \subset T^n a$ and $\mu(T^{-1}a) = \mu(a)$.

3 Transfer operator and characteristic function operator

We continue setting up the theory for Gibbs-Markov system (Ω, μ, T, α) by introducing its transfer operators. Since $T\alpha \subset \sigma(\alpha)$, it follows that for every $n \in \mathbb{N}$, $T^n(\bigvee_{i=0}^{n-1} T^{-i}\alpha) = T\alpha$. Fix a partition β (which may be coarser than α) such that

$$\sigma(T\alpha) = \sigma(\beta). \quad (2)$$

The Perron-Frobenius transfer operator $\mathcal{L}_T : L^1(\mu) \rightarrow L^1(\mu)$ is defined by:

$$\mathcal{L}_T f := \sum_{b \in \beta} \mathbf{1}_b \sum_{a \in \alpha, Ta \supset b} v'_a \cdot f \circ v_a,$$

and the transfer operator for T^n is hence of the form

$$\mathcal{L}_{T^n} f = \sum_{b \in \beta} \mathbf{1}_b \sum_{a \in \bigvee_{i=0}^{n-1} T^{-i}\alpha, T^n a \supset b} v'_a \cdot f \circ v_a.$$

\mathcal{L}_T satisfies and is uniquely characterized by:

$$\int_\Omega \mathcal{L}_T f \cdot g d\mu = \int_\Omega f \cdot g \circ T d\mu, \quad \forall g \in L^\infty(\mu).$$

It can be easily derived from the above equation that

$$\mathcal{L}_{T^n} = \mathcal{L}_T^n$$

and since μ is T -invariant

$$\mathcal{L}_T \mathbf{1} = \mathbf{1}. \quad (3)$$

We will use \mathcal{L} for \mathcal{L}_T when T is fixed.

Given a measurable function $f : \Omega \rightarrow \mathbb{R}$ and $t \in \mathbb{R}$, the characteristic function operator $\mathcal{L}_{f,t} : L^1(\mu) \rightarrow L^1(\mu)$ is defined as:

$$\mathcal{L}_{f,t}g := \mathcal{L}(e^{itf} \cdot g), \quad \forall g \in L^1(\mu).$$

Then

$$\mathcal{L}_{f,0} = \mathcal{L}.$$

Note that $D_\alpha(\cdot) \leq D_\beta(\cdot) \leq D_\Omega(\cdot) \leq \max\{\frac{2\|\cdot\|_\infty}{r}, D_\alpha(\cdot)\}$, hence the functions in L_β^∞ have finite D_α norms, and the norm $\|\cdot\|_{\infty,\beta}$ is equivalent to the norm $\|\cdot\|_{\infty,\alpha}$ and to the norm $\|\cdot\|_{\infty,\Omega}$. From now on, we write for simplicity

$$L := L_\beta^\infty, \quad \|\cdot\| := \|\cdot\|_{\infty,\beta}.$$

As no confusion should appear, we use the same notation $\|\cdot\|$ for the operator norm on L . It is not hard to see that \mathcal{L} and $\mathcal{L}_{f,t}$ are both bounded linear operators on L . In fact, we have the following estimates.

Lemma 3.1 ([2, Proposition 2.1, Theorem 2.4]). *There exist constants M and M_1 such that for any $f, g \in L$ and $s, t \in \mathbb{R}$, we have*

$$\|\mathcal{L}_{f,t}^n g\| \leq (M + M_1 D_\alpha e^{itf})(r^n D_\beta g + \|g\|_1) \quad (4)$$

and

$$\|\mathcal{L}_{f,t} - \mathcal{L}_{f,s}\| \leq (M + M_1 |s| D_\alpha f)(\|e^{i(t-s)f} - \mathbf{1}\|_1 + |t - s| D_\alpha f).$$

In particular, let $s = 0$, we have for every $t \in \mathbb{R}$ and $f \in L$,

$$\|\mathcal{L}_{f,t} - \mathcal{L}\| \leq 2M|t| \cdot \|f\|.$$

The following lemma, adapted from [23, Proposition 3], provides the Taylor expansion of $\mathcal{L}_{f,t}$ around \mathcal{L} .

Lemma 3.2. *For any $f \in L$ and $t \in \mathbb{R}$, there exist bounded linear operators $\mathcal{L}_f^{(n)}$ and $\mathcal{L}_{f,t}^{(n)}$ on L such that*

$$\mathcal{L}_{f,t} = \mathcal{L} + \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathcal{L}_f^{(n)}$$

converges absolutely with $\|\mathcal{L}_f^{(n)}\| \leq \|\mathcal{L}\| \|f\|^n$, and for every $m \in \mathbb{N}$,

$$\mathcal{L}_{f,t} = \mathcal{L} + t\mathcal{L}_f^{(1)} + \cdots + \frac{t^{m-1}}{(m-1)!} \mathcal{L}_f^{(m-1)} + \mathcal{L}_{f,t}^{(m)} \quad (5)$$

with $\|\mathcal{L}_{f,t}^{(m)}\| \leq \|\mathcal{L}\| |t|^m \|f\|^m e^{|t| \|f\|}$.

Remark 3.3. $\mathcal{L}_f^{(n)}$ are just derivatives of $\mathcal{L}_{f,t}$ around $t = 0$ when f is fixed.

Proof. Since for any $f, g \in L$,

$$\|f \cdot g\|_\infty \leq \|f\|_\infty \cdot \|g\|_\infty \text{ and}$$

$$D_\beta(f \cdot g) \leq D_\beta(f) \cdot \|g\|_\infty + \|f\|_\infty \cdot D_\beta(g),$$

we have

$$\|f \cdot g\| \leq \|f\| \cdot \|g\|. \quad (6)$$

Therefore $\|f^n\| \leq \|f\|^n$, and

$$\left\| \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mathcal{L}(f^n g) \right\| \leq \sum_{n=0}^{\infty} \frac{1}{n!} \|\mathcal{L}\| \|f\|^n |t|^n \|g\|$$

converges absolutely. Let $\mathcal{L}_f^{(n)}(\cdot) := i^n \mathcal{L}(f^n \cdot)$ and the expansion follows. \square

One of the underlying tools throughout this paper is the spectral gap property of the transfer operator \mathcal{L} , which has been proved in [2, Theorem 1.6]. We explain it briefly here. Because that L -bounded sets are precompact in L^1 and because of Lemma 3.1 and a theorem of Ionescu-Tulcea and Marinescu ([20]), \mathcal{L} is quasi-compact on L . Notice that 1 is an eigenvalue of \mathcal{L} and is a maximal eigenvalue of \mathcal{L} on L by (3) since \mathcal{L} contracts L^∞ . Also it is known that a (topologically) mixing Gibbs-Markov map is exact ([3, 2]) and hence is strong-mixing when it is measure-preserving. So 1 is the unique maximal eigenvalue and is simple. Hence the transfer operator \mathcal{L} of a mixing Gibbs-Markov map has the spectral gap property on L , namely one can decompose \mathcal{L} on L as

$$\mathcal{L} = P + N \quad (7)$$

so that $Pf = \int_\Omega f d\mu$, $PN = NP = 0$ and $\mathfrak{r}(N) < 1$, where $\mathfrak{r}(\cdot)$ is spectral radius. P is the eigenprojection of \mathcal{L} with respect to the eigenvalue 1 and the spectrum of N is all the remaining spectrum of \mathcal{L} . For any complex number z not in the spectrum of \mathcal{L} , denote by $R(z; \mathcal{L})$ the resolvent of \mathcal{L} , $(z\mathcal{I} - \mathcal{L})^{-1}$. According to [13, VII], one can calculate P and N by integrating the product of the resolvent by suitable functions analytic on neighborhoods of the spectrum of \mathcal{L} . In fact,

$$P = \frac{1}{2\pi i} \int_{C_1} R(z; \mathcal{L}) dz, \quad N = \frac{1}{2\pi i} \int_{C_2} z R(z; \mathcal{L}) dz \quad (8)$$

where C_1 is a small circle around 1 of radius $\frac{1-\mathfrak{r}(N)}{3}$ and C_2 is a circle around 0 of radius $\frac{1+2\mathfrak{r}(N)}{3}$ so that C_1 and C_2 are disjoint and that the spectrum of \mathcal{L} except for the eigenvalue 1 is totally contained within C_2 . For every positive integer k ,

$$N^k = \frac{1}{2\pi i} \int_{C_2} z^k R(z; \mathcal{L}) dz.$$

We will call

$$\rho_1 := \frac{1 - \mathfrak{r}(N)}{3}, \quad \rho_2 := \frac{1 + 2\mathfrak{r}(N)}{3}.$$

By perturbation theory, the characteristic function operators also satisfy the above properties.

Lemma 3.4. *There exists a real number $a > 0$ such that if $|t| \cdot \|f\| < a$ then $\mathcal{L}_{f,t}$ has the spectral gap property on L with the decomposition:*

$$\mathcal{L}_{f,t} = \lambda_{f,t} P_{f,t} + N_{f,t}$$

where

1. $\lambda_{f,t}$ is the unique eigenvalue of largest modulus of $\mathcal{L}_{f,t}$, $\lambda_{f,t}$ is a simple eigenvalue and $|\lambda_{f,t}| \in (1 - \rho_1, 1 + \rho_1)$;
2. $P_{f,t}$ is the eigenprojection of $\mathcal{L}_{f,t}$ with respect to $\lambda_{f,t}$, in the form

$$P_{f,t} = \frac{1}{2\pi i} \int_{C_1} R(z; \mathcal{L}_{f,t}) dz;$$

3. $\mathfrak{r}(N_{f,t}) < \rho_2$ and $P_{f,t} N_{f,t} = N_{f,t} P_{f,t} = 0$, in fact,

$$N_{f,t} = \frac{1}{2\pi i} \int_{C_2} z R(z; \mathcal{L}_{f,t}) dz;$$

4. fix an $f \in L$, then $t \mapsto \lambda_{f,t}$, $t \mapsto P_{f,t}$ and $t \mapsto N_{f,t}$ are analytic on $(-a/\|f\|, a/\|f\|)$.

Here C_1, C_2 are the same circles as in (8).

This lemma is essentially [23, Proposition 4]. For our purpose we take expansions of the operators to higher orders in the next lemma.

Lemma 3.5. *There exist constants $M > 0$ and $0 < a < M^{-1}$ such that if $|t| \cdot \|f\| < a$ is small enough, then*

1. $P_{f,t}$ has an expansion

$$P_{f,t} = P + tP_f^{(1)} + \frac{t^2}{2}P_f^{(2)} + P_{f,t}^{(3)},$$

where $\|P_f^{(i)}\| \leq M\|f\|^i$, for $i = 1, 2$, and $\|P_{f,t}^{(3)}\| \leq M|t|^3\|f\|^3e^{3|t|\|f\|}$;

2. similarly, $\lambda_{f,t}$ expands as

$$\lambda_{f,t} = 1 + t\lambda_f^{(1)} + \frac{t^2}{2}\lambda_f^{(2)} + \lambda_{f,t}^{(3)},$$

where $|\lambda_f^{(i)}| \leq M\|f\|^i$, for $i = 1, 2$, and $|\lambda_{f,t}^{(3)}| \leq M|t|^3\|f\|^3e^{3|t|\|f\|}$;

3. for all $n \in \mathbb{N}$,

$$\|N_{f,t}^n \mathbf{1}\| \leq \rho_2^n \frac{M|t|\|f\|}{1 - M|t|\|f\|}.$$

Proof. We use notations from Lemma 3.4. Notice that for any z in the resolvent set of \mathcal{L} , if $|t| \cdot \|f\|$ is so small that $\|\mathcal{L}_{f,t} - \mathcal{L}\| \cdot \|R(z, \mathcal{L})\| < 1$ then z is also in the resolvent set of $\mathcal{L}_{f,t}$ and

$$R(z; \mathcal{L}_{f,t}) = R(z; \mathcal{L}) \sum_{n=0}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R(z; \mathcal{L}))^n \quad (9)$$

converges absolutely.

1. We use the resolvent equation (9) to calculate $P_{f,t}$ as follows. Choose a small enough such that $\|\mathcal{L}_{f,t} - \mathcal{L}\| \cdot \sup_{z \in C_1} \|R(z, \mathcal{L})\| < 1$ whenever $|t|\|f\| < a$. Then, denoting by $R := R(z; \mathcal{L})$,

$$\begin{aligned} P_{f,t} &= \frac{1}{2\pi i} \int_{C_1} R(z; \mathcal{L}_{f,t}) dz \\ &= \frac{1}{2\pi i} \int_{C_1} R(z; \mathcal{L}) \sum_{n=0}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R(z; \mathcal{L}))^n dz \\ &= P + \frac{1}{2\pi i} \int_{C_1} R(z; \mathcal{L}) \sum_{n=1}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R(z; \mathcal{L}))^n dz \\ &\stackrel{(5)}{=} P + t \frac{1}{2\pi i} \int_{C_1} R \mathcal{L}_f^{(1)} R dz + \frac{t^2}{2} \frac{1}{2\pi i} \int_{C_1} \left\{ R \mathcal{L}_f^{(2)} R + 2R \left(\mathcal{L}_f^{(1)} R \right)^2 \right\} dz \\ &\quad + \frac{1}{2\pi i} \int_{C_1} \left\{ t R \mathcal{L}_{f,t}^{(2)} R \mathcal{L}_f^{(1)} R + t R \mathcal{L}_f^{(1)} R \mathcal{L}_{f,t}^{(2)} R + R \left(\mathcal{L}_{f,t}^{(2)} R \right)^2 \right. \\ &\quad \left. + R \sum_{n=3}^{\infty} \left(\mathcal{L}_{f,t}^{(1)} R \right)^n \right\} dz. \end{aligned}$$

Define corresponding operators to write the last equation in the form

$$P_{f,t} = P + tP_f^{(1)} + \frac{t^2}{2}P_f^{(2)} + P_{f,t}^{(3)}.$$

Then there exists constants M_1 and M_2 such that when $|t|\|f\|$ is small,

$$\begin{aligned} \|P_f^{(1)}\| &\leq M_2 \|\mathcal{L}_f^{(1)}\| \leq M_1 \|\mathcal{L}\| \|f\|, \\ \|P_f^{(2)}\| &\leq M_2 (\|\mathcal{L}_f^{(2)}\| + 2\|\mathcal{L}_f^{(1)}\|^2) \leq M_1 \|\mathcal{L}\|^2 \|f\|^2, \\ \|P_{f,t}^{(3)}\| &\leq M_2 (|t| \cdot \|\mathcal{L}_{f,t}^{(2)}\| \cdot \|\mathcal{L}_f^{(1)}\| + \|\mathcal{L}_{f,t}^{(2)}\|^2 + \sum_{n=3}^{\infty} M_2^n \|\mathcal{L}_{f,t}^{(1)}\|^n) \\ &\leq M_1 |t|^3 \|f\|^3 e^{3|t|\|f\|}. \end{aligned}$$

2. Let B be the Banach space of all bounded linear operators from L to itself. Take a linear functional $\varphi \in B^*$ such that $\|\varphi\|_{B^*} = 1$ and $\varphi(P) = 1$. Then because

$$\mathcal{L}_{f,t} P_{f,t} = \lambda_{f,t} P_{f,t},$$

we have

$$\lambda_{f,t} = \frac{\varphi(\mathcal{L}_{f,t}P_{f,t})}{\varphi(P_{f,t})}.$$

Define $P_{f,t}^{(1)} = tP_f^{(1)} + \frac{t^2}{2}P_f^{(2)} + P_f^{(3)}$, then $\|P_{f,t}^{(1)}\| \leq M_1|t|\|f\|e^{|t|\|f\|}$. Choose a small such that $ae^a < M_1^{-1}$ then when $|t|\|f\| < a$,

$$\frac{1}{\varphi(P_{f,t})} = \frac{1}{1 + \varphi(P_{f,t}^{(1)})} = \sum_{n=0}^{\infty} (-1)^n \varphi(P_{f,t}^{(1)})^n$$

since $\varphi(P_{f,t}^{(1)}) \leq M_1|t|\|f\|e^{|t|\|f\|}$. Hence the expansions of $\mathcal{L}_{f,t}$ and $P_{f,t}$ lead to the expansion of $\lambda_{f,t}$.

3. The same resolvent equation is used in the calculation of $N_{f,t}$. Choose a small enough such that $\|\mathcal{L}_{f,t} - \mathcal{L}\| \cdot \sup_{z \in C_2} (\|R(z, \mathcal{L})\| + \|R(z, \mathcal{L})\|^2) < 1$ whenever $|t|\|f\| < a$, then

$$\begin{aligned} N_{f,t}^n &= \frac{1}{2\pi i} \int_{C_2} z^n R(z; \mathcal{L}_{f,t}) dz \\ &= \frac{1}{2\pi i} \int_{C_2} z^n R(z; \mathcal{L}) \sum_{m=0}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R(z; \mathcal{L}))^m dz \\ &= N + \frac{1}{2\pi i} \int_{C_2} z^n R(z; \mathcal{L}) \sum_{m=1}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R(z; \mathcal{L}))^m dz. \end{aligned}$$

Notice that $N\mathbf{1} = 0$, whence we have:

$$\begin{aligned} \|N_{f,t}^n \mathbf{1}\| &\leq \frac{1}{2\pi} \left\| \int_{C_2} z^n R(z; \mathcal{L}) \sum_{m=1}^{\infty} ((\mathcal{L}_{f,t} - \mathcal{L})R(z; \mathcal{L}))^m dz \right\| \\ &\leq \rho_2^n \frac{M_2|t|\|f\|}{1 - M_2|t|\|f\|} \end{aligned}$$

for some constant M_2 with $a < M_2^{-1}$.

□

When f is fixed, since $\lambda_{f,t}$ is analytic with respect to t around 0 (Lemma 3.4), the coefficients $\lambda_f^{(i)}$ in the expansion of $\lambda_{f,t}$ are just the corresponding derivatives of $\lambda_{f,t}$ with respect to t at $t = 0$. They can be calculated in the following manner.

Lemma 3.6 ([23, Lemmes 2, 3, 6]). *Let $f \in L$ with $\int_{\Omega} f d\mu = 0$. Then*

$$\lambda_f^{(1)} = 0, \quad \lambda_f^{(2)} = - \lim_{n \rightarrow \infty} \frac{1}{n} \int_{\Omega} \left(\sum_{m=0}^{n-1} f \circ T^m \right)^2 d\mu.$$

The limit exists, and $\lambda_f^{(2)} \neq 0$ if and only if f is not of the form $\varphi \circ T - \varphi$ for any $\varphi \in L$.

Proof. One only needs to notice that the transfer operator and its spectral decomposition in [23] share the same spectral properties and expand in the same way as in our settings when f is fixed, to which the proofs of [23, Lemmes 2, 3, 6] refer, hence the proofs carry over to our settings. \square

The asymptotic variance of f will be denoted by

$$\sigma_f^2 := \lim_{n \rightarrow \infty} \frac{1}{n} \int_{\Omega} \left(\sum_{m=0}^{n-1} f \circ T^m \right)^2 d\mu.$$

4 Central limit theorem for arrays

As mentioned in the introduction, in this section we prove a central limit theorem based on the tools developed in Section 3 for arrays of functions of class L in Gibbs-Markov systems. The following theorem appeared as part of the third author's PhD thesis [26].

Theorem 4.1. *Consider a Gibbs-Markov system (Ω, μ, T, α) , a sequence $\{f_n\} \subset L$ with $\int_{\Omega} f_n d\mu = 0$ and not of the form $\varphi \circ T - \varphi$ for any $\varphi \in L$ and a sequence of positive integers $k_n \rightarrow \infty$. If*

$$\lim_{n \rightarrow \infty} \frac{\|f_n\|^3}{\sqrt{k_n} \sigma_n^3} = 0,$$

where $\sigma_n^2 := \sigma_{f_n}^2$ is the asymptotic variance of f_n , then

$$\frac{f_n + f_n \circ T + \dots + f_n \circ T^{k_n-1}}{\sqrt{k_n} \sigma_n^2}$$

converges in distribution to the standard normal law $\mathcal{N}(0, 1)$.

Proof. Let $S_n = f_n + f_n \circ T + \dots + f_n \circ T^{k_n-1}$. It can be easily verified that

$$\mathcal{L}_{f_n, t}^{k_n} \mathbf{1} = \mathcal{L}^{k_n} e^{it S_n}.$$

Let $t_n = \frac{t}{\sqrt{k_n} \sigma_n}$, then we have

$$\begin{aligned} \int_{\Omega} e^{it \frac{S_n}{\sqrt{k_n} \sigma_n}} d\mu &= \int_{\Omega} e^{it_n S_n} d\mu = \int_{\Omega} \mathcal{L}^{k_n} e^{it_n S_n} d\mu \\ &= \int_{\Omega} \mathcal{L}_{f_n, t_n}^{k_n} \mathbf{1} d\mu. \end{aligned}$$

Since by assumption as $n \rightarrow \infty$,

$$|t_n| \|f_n\| = |t| \frac{\|f_n\|}{\sqrt{k_n} \sigma_n} \rightarrow 0,$$

when n is large, according to Lemma 3.4

$$\int_{\Omega} e^{it \frac{S_n}{\sqrt{k_n \sigma_n}}} d\mu = \lambda_{f_n, t_n}^{k_n} \int_{\Omega} P_{f_n, t_n} \mathbf{1} d\mu + \int_{\Omega} N_{f_n, t_n}^{k_n} \mathbf{1} d\mu.$$

Lemma 3.5 implies that

$$\lim_{n \rightarrow \infty} \int_{\Omega} P_{f_n, t_n} \mathbf{1} d\mu = \int_{\Omega} P \mathbf{1} d\mu = 1$$

and for some constant M ,

$$\left| \int_{\Omega} N_{f_n, t_n}^{k_n} \mathbf{1} d\mu \right| \leq \rho_2^{k_n} \frac{M |t_n| \|f_n\|}{1 - M |t_n| \|f_n\|} \rightarrow 0.$$

By Lemma 3.5 and Lemma 3.6, when n is large,

$$\begin{aligned} \lambda_{f_n, t_n} &= 1 - \frac{1}{2} \sigma_n^2 t_n^2 + \lambda_{f_n, t_n}^{(3)} \\ &= 1 - \frac{t^2}{2k_n} + \lambda_{f_n, t_n}^{(3)}, \end{aligned}$$

where $|\lambda_{f_n, t_n}^{(3)}| \leq M |t_n|^3 \|f_n\|^3 e^{3|t_n| \|f_n\|}$. Hence the assumption that

$$k_n |t_n|^3 \|f_n\|^3 = t^3 \frac{\|f_n\|^3}{\sqrt{k_n \sigma_n^3}} \rightarrow 0$$

implies

$$\lambda_{f_n, t_n}^{k_n} \rightarrow e^{-\frac{t^2}{2}}.$$

This shows

$$\lim_{n \rightarrow \infty} \int_{\Omega} e^{it \frac{S_n}{\sqrt{k_n \sigma_n}}} d\mu = e^{-\frac{1}{2} t^2},$$

finishing the proof of the theorem. \square

Corollary 4.2. *Let (Ω, μ, T, α) be a Gibbs-Markov map. Then every function*

$$f = \sum_{n=1}^{\infty} \gamma_n g_n$$

with $g_n \in L$, $\gamma_n \in \mathbb{R}$, $\int_{\Omega} g_n d\mu = 0$, $\sup_{n \in \mathbb{N}} \|g_n\|_2 < \infty$, $\sup_{n \in \mathbb{N}} |\gamma_n| \|g_n\| < \infty$ and $\sum_{k=n+1}^{\infty} |\gamma_k| \leq K n^{-3-\eta}$ (for some constants $K, \eta > 0$) satisfies the central limit theorem in the form

$$\frac{1}{\sqrt{n} \sigma_n} \sum_{j=0}^{n-1} f \circ T^j \Rightarrow \mathcal{N}(0, 1)$$

for some sequence $\sigma_n > 0$, provided the asymptotic variances of $\sum_{k=1}^n \gamma_k g_k$ are bounded away from 0 uniformly.

Remark 4.3. If the asymptotic variances in the previous corollary converge to zero, then the ergodic sums normed by \sqrt{n} converge to 0 stochastically.

Proof. Let $l_n \in \mathbb{N}$ satisfy

$$\lim_{n \rightarrow \infty} n^{-1} l_n^6 = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} l_n^{-6-2\eta} n = 0.$$

Define

$$f_n = \sum_{j=1}^{l_n} \gamma_j g_j$$

and denote by σ_n^2 the asymptotic variance of f_n . Note that $f_n \in L$ with $\|f_n\| \leq \sum_{k=1}^{l_n} |\gamma_k| \|g_k\| \leq C l_n$ for some constant $C > 0$ and hence (since $\inf \sigma_n > 0$)

$$\frac{\|f_n\|^3}{\sqrt{n} \sigma_n^3} = O\left(\frac{l_n^3}{\sqrt{n}}\right) \rightarrow 0.$$

Take $k_n = n$ in Theorem 4.1 to deduce that

$$\frac{f_n + f_n \circ T + \dots + f_n \circ T^{n-1}}{\sqrt{n} \sigma_n}$$

converges to the standard normal distribution.

Now by Chebychev's inequality with $M = \sup_{n \in \mathbb{N}} \|g_n\|_2$, for any $\epsilon > 0$

$$\begin{aligned} & \mu \left(\left\{ x \in \Omega : \left| \sum_{j=0}^{n-1} \sum_{k=l_n+1}^{\infty} \gamma_k g_k(T^j(x)) \right| \geq \epsilon \sqrt{n} \sigma_n \right\} \right) \\ & \leq \frac{1}{\epsilon^2 n \sigma_n^2} \int_{\Omega} \left(\sum_{j=0}^{n-1} \sum_{k=l_n+1}^{\infty} \gamma_k g_k \circ T^j \right)^2 d\mu \\ & \leq \frac{1}{\epsilon^2 n \sigma_n^2} n^2 M^2 \left(\sum_{k=l_n+1}^{\infty} |\gamma_k| \right)^2 \\ & \leq O(n(l_n^{-3-\eta})^2), \end{aligned}$$

which converges to 0. It follows that $\frac{1}{\sqrt{n} \sigma_n} \sum_{j=0}^{n-1} f_n \circ T^j$ and $\frac{1}{\sqrt{n} \sigma_n} \sum_{j=0}^{n-1} f \circ T^j$ have the same limiting distribution, whence the corollary. \square

Example 4.4. Let $(\Omega, \mathcal{B}, \mu, T, \alpha)$ denote the continued fraction transformation with $\Omega = (0, 1)$ and μ the Gauss measure. For every irrational $x \in (0, 1)$, denote by $(x_n)_{n \in \mathbb{N}}$ its continued fraction expansion. Let $a_n := \{x : x_0 = n\}$ for every $n \in \mathbb{N}$, then the partition $\alpha = \{a_n : n \in \mathbb{N}\}$. Let $\eta \in (0, \frac{1}{2})$, define

$$\begin{aligned} m_n &:= \lfloor -\log_r n^2 \rfloor, \quad \ell_n := r^{-m_n}, \\ \gamma_n &:= r^{(2+\eta)m_n}, \quad g_n := \ell_n \mathbf{1}_{a_{\lfloor \ell_n \rfloor}} \circ T^{m_n} - \ell_n \mu(a_{\lfloor \ell_n \rfloor}). \end{aligned}$$

Here we denote by $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ the usual floor and ceiling functions for real numbers. Recall that $r \in (0, 1)$ is the constant in (1). In current case of continued fraction transformation, we can take $r = 2/3$ (cf. [2, Example 2]). It is easy to see that $g_n \in L$ and $\int_{\Omega} g_n d\mu = 0$.

$$\begin{aligned}\|g_n\|_{\infty} &= \ell_n(1 - \mu(a_{\lfloor \ell_n \rfloor})), \quad D_{\Omega} g_n = \ell_n r^{-m_n} = r^{-2m_n} \asymp n^4, \\ |\gamma_n| \|g_n\| &= O(1) r^{\eta m_n} = O(1) n^{-2\eta}, \\ \|g_n\|_2 &= \ell_n \sqrt{\mu(a_{\lfloor \ell_n \rfloor})(1 - \mu(a_{\lfloor \ell_n \rfloor}))} = O(1) \ell_n / \sqrt{(\ell_n(\ell_n + 1))},\end{aligned}$$

$$\begin{aligned}\sum_{k=n}^{\infty} |\gamma_k| &\leq \sum_{d=m_n}^{\infty} \sum_{\lceil r^{-d/2} \rceil \leq k \leq \lfloor r^{-(d+1)/2} \rfloor} r^{(2+\eta)d} \\ &\leq \sum_{d=m_n}^{\infty} r^{(2+\eta)d} (r^{-(d+1)/2} - r^{-d/2}) \leq O(n^{-3-2\eta}).\end{aligned}$$

It follows that $f = \sum_{j=1}^{\infty} \gamma_j g_j$ satisfies the assumptions in the corollary, hence the central limit theorem holds:

$$\frac{1}{\sqrt{n}\sigma_n} \sum_{j=0}^{n-1} f \circ T^j \Rightarrow \mathcal{N}(0, 1).$$

We remark that $f \in L^2(\mu)$ but $f \notin L$. In fact, for large n ,

$$\begin{aligned}D_{\alpha} f &\geq D_{\bigvee_{i=0}^{m_n-1} T^{-i}\alpha} f \geq r^{-m_n} \sum_{d \geq m_n} \sum_{\lceil r^{-d/2} \rceil \leq k < \lfloor r^{-(d+1)/2} \rfloor} r^{(2+\eta)d} \cdot r^{-d} \\ &= r^{-m_n} O(r^{(\frac{1}{2}+\eta)m_n}) = O(r^{(-\frac{1}{2}+\eta)m_n}) = O(n^{1-2\eta}),\end{aligned}$$

hence $D_{\alpha} f$ is infinite. This calculation also indicates that for any $a \in \bigvee_{i=0}^{n-1} T^{-i}\alpha$, $D_a f$ is infinite.

We wish to point out that the proofs of the CLT presented in this section were detailed in the context of Gibbs-Markov maps but also hold in other, more general settings of mixing dynamical systems. The main technique is the spectral gap property of Ionescu-Tulcea and Marinescu [20] which allows for the decomposition in equation (7). This property holds in all generality for maps which satisfy a Doeblin-Fortet inequality as in Lemma 3.1. Our Lemmas 3.4, 3.5 and 3.6 are instrumental in proving the CLT of Theorem 4.1 and so this CLT holds in *any* setting in which the above mentioned Lemmas are valid.

Note that identifying the context of a Banach space of functions along with a pair of norms satisfying our assumptions is a delicate but necessary task, without which the theorem lacks relevant examples, and we refrain from formulating our theorems in such an abstract albeit empty context. Other known examples of general settings to which Theorem 4.1 applies, beyond the Gibbs-Markov systems presented in Section 2, include maps of the interval endowed with the bounded variation norm, as well as Young towers endowed with the Hölder norm.

5 A CLT after Lindeberg

We prove a CLT for a dynamical array, and later apply it to Birkhoff sums. The notion of dynamical array is also considered in [12].

Definition 5.1. A dynamical array is a sequence $\{(F_{n,i}, \tau_{n,i}) : i = 1, \dots, k_n \in \mathbb{N}\}_{n \in \mathbb{N}}$ where $F_{n,i}$ are real valued functions defined on a dynamical system (Ω, T) of form

$$F_{n,i} = \sum_{l=1}^{n_i} f_{n,i,l} \circ T^{l-1}, \quad i = 1, \dots, k_n$$

and where $\tau_{n,i-1} + n_{i-1} \leq \tau_{n,i} \in \mathbb{N}$.

We recall some notations. Let (Ω, μ, T, α) be a mixing Gibbs-Markov system. β is a partition of Ω satisfying (2), $\sigma(\beta) = \sigma(T\alpha)$. A real number r inducing a metric on Ω is given in Definition 2.1. The transfer operator \mathcal{L} has the decomposition (7) on $L = L_\beta^\infty$, i.e

$$\mathcal{L}f = \int_\Omega f d\mu + Nf \quad (10)$$

for all $f \in L$. Let $\rho := \mathfrak{r}(N) \in (0, 1)$.

Remark 5.2. It is known ([18, Corollaire 1]) that the essential spectrum of \mathcal{L} is at most r due to Lemma 3.1, but the relation between ρ and r is unclear to us.

Theorem 5.3. Let $\{(F_{n,i}, \tau_{n,i}) : i = 1, \dots, k_n \in \mathbb{N}\}_{n \in \mathbb{N}}$ be a dynamical array defined on a mixing Gibbs-Markov system (Ω, μ, T, α) with $F_{n,i} = \sum_{l=1}^{n_i} f_{n,i,l} \circ T^{l-1}$. Suppose that every $f_{n,i,l} \in L$ is centered, i.e. $\int_\Omega f_{n,i,l} d\mu = 0$. Let

$$\check{s}_n^2 := \text{Var}(F_{n,1} \circ T^{\tau_{n,1}} + F_{n,2} \circ T^{\tau_{n,2}} + \dots + F_{n,k_n} \circ T^{\tau_{n,k_n}}),$$

and

$$m_n := \inf_{2 \leq i \leq k_n} \tau_{n,i} - \tau_{n,i-1} - n_{i-1}.$$

Assume the following properties for this array.

1. For every $n \in \mathbb{N}$,

$$\check{s}_n \neq 0.$$

2.

$$\limsup_{n \rightarrow \infty} k_n^2 \rho^{m_n} < \infty. \quad (11)$$

3.

$$\limsup_{n \rightarrow \infty} \rho^{m_n} \sum_{1 \leq i \leq k_n} r^{n_i} \frac{\|F_{n,i}\|}{\check{s}_n} < \infty. \quad (12)$$

4. The Lindeberg condition holds, i.e. for every $\epsilon > 0$

$$L_{n,\epsilon} := \frac{1}{\check{s}_n^2} \sum_{i=1}^{k_n} \int_{\Omega} F_{n,i}^2 \cdot \mathbf{1}_{\{|F_{n,i}| \geq \epsilon \check{s}_n\}} d\mu \xrightarrow{n \rightarrow \infty} 0. \quad (13)$$

Then, this array satisfies a CLT, i.e.

$$\frac{F_{n,1} \circ T^{\tau_{n,1}} + F_{n,2} \circ T^{\tau_{n,2}} + \dots + F_{n,k_n} \circ T^{\tau_{n,k_n}}}{\check{s}_n} \Rightarrow \mathcal{N}(0, 1).$$

Remark 5.4. Condition (11) implies that m_n must tend to ∞ when $k_n \rightarrow \infty$.

Lemma 5.5. Under the same assumptions as in Theorem 5.3,

1. There exists a constant C independent of n such that with

$$\rho_n := \rho^{m_n}, \quad r_n := r^{m_n},$$

for every $f \in L$, $n \in \mathbb{N}$ and $1 \leq i < j \leq k_n$,

$$\|N^{\tau_{n,j}-\tau_{n,i}} f\| \leq C\rho_n \|f\|_1 + C\rho_n r_n^{j-i-1} r^{n_i+\dots+n_{j-1}} D_{\beta} f, \quad (14)$$

$$\|N^{\tau_{n,j}} f\| \leq C\rho_n \|f\|_1 + C\rho_n r^{\tau_{n,j-1}+n_{j-1}} D_{\beta} f; \quad (15)$$

2. the array is asymptotically negligible,

$$\lim_{n \rightarrow \infty} \sup_{1 \leq i \leq k_n} \int_{\Omega} \frac{F_{n,i}^2}{\check{s}_n^2} d\mu = 0; \quad (16)$$

3. an asymptotic variance formula holds,

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \int_{\Omega} \frac{F_{n,i}^2}{\check{s}_n^2} d\mu = 1. \quad (17)$$

Proof. 1. Because $\mathcal{L} = P + N$ with $PN = NP = 0$,

$$N^{\tau_{n,j}-\tau_{n,i}} f = N^{\tau_{n,i+1}-\tau_{n,i}-n_i} \mathcal{L}^{\tau_{n,j}-\tau_{n,i+1}+n_i} f.$$

Therefore, by (4) (with $t = 0$), we have

$$\begin{aligned} \|N^{\tau_{n,j}-\tau_{n,i}} f\| &\leq \|N^{\tau_{n,i+1}-\tau_{n,i}-n_i}\| \|\mathcal{L}^{\tau_{n,j}-\tau_{n,i+1}+n_i} f\| \\ &\leq O(1) \rho^{\tau_{n,i+1}-\tau_{n,i}-n_i} (\|f\|_1 + r^{\tau_{n,j}-\tau_{n,i+1}+n_i} D_{\beta} f) \\ &\leq O(1) \rho^{m_n} (\|f\|_1 + r^{(j-i-1)m_n} r^{n_i+\dots+n_{j-1}} D_{\beta} f). \end{aligned}$$

Similarly, for $2 \leq j \leq k_n$, (15) follows from

$$N^{\tau_{n,j}} f = N^{\tau_{n,j}-\tau_{n,j-1}-n_{j-1}} \mathcal{L}^{\tau_{n,j-1}+n_{j-1}} f$$

2. This is implied by the Lindeberg condition (13), for

$$\begin{aligned} \int_{\Omega} \frac{F_{n,i}^2}{\check{s}_n^2} d\mu &\leq \int_{\Omega} \frac{F_{n,i}^2}{\check{s}_n^2} \cdot \mathbf{1}_{\{|\frac{F_{n,i}}{\check{s}_n}| < \epsilon\}} d\mu + \sum_{i=1}^{k_n} \int_{\Omega} \frac{F_{n,i}^2}{\check{s}_n^2} \cdot \mathbf{1}_{\{|\frac{F_{n,i}}{\check{s}_n}| \geq \epsilon\}} d\mu \\ &\leq \epsilon^2 + L_{n,\epsilon}. \end{aligned} \quad (18)$$

3. Use the transfer operator \mathcal{L} to expand the total variance

$$\begin{aligned} \check{s}_n^2 &= \int_{\Omega} (F_{n,1} \circ T^{\tau_{n,1}} + \cdots + F_{n,k_n} \circ T^{\tau_{n,k_n}})^2 d\mu \\ &= \sum_{i=1}^{k_n} \int_{\Omega} F_{n,i}^2 d\mu + 2 \sum_{1 \leq i < j \leq k_n} \int_{\Omega} F_{n,i} \cdot F_{n,j} \circ T_n^{\tau_{n,j} - \tau_{n,i}} d\mu \\ &= \sum_{i=1}^{k_n} \int_{\Omega} F_{n,i}^2 d\mu + 2 \sum_{1 \leq i < j \leq k_n} \int_{\Omega} \mathcal{L}^{\tau_{n,j} - \tau_{n,i}} F_{n,i} \cdot F_{n,j} d\mu \\ &\stackrel{(10)}{=} \sum_{i=1}^{k_n} \int_{\Omega} F_{n,i}^2 d\mu + 2 \sum_{1 \leq i < j \leq k_n} \int_{\Omega} N^{\tau_{n,j} - \tau_{n,i}} F_{n,i} \cdot F_{n,j} d\mu. \end{aligned}$$

The last equality holds because $\int_{\Omega} F_{n,i} d\mu = 0$. Estimate

$$\begin{aligned} &\left| \frac{1}{\check{s}_n^2} \sum_{1 \leq i < j \leq k_n} \int_{\Omega} N^{\tau_{n,j} - \tau_{n,i}} F_{n,i} \cdot F_{n,j} d\mu \right| \\ &\stackrel{(14)}{\leq} \frac{1}{\check{s}_n^2} C \sum_{1 \leq i < j \leq k_n} (\rho_n \|F_{n,i}\|_1 + \rho_n r_n^{j-i-1} r^{n_i + \cdots + n_{j-1}} D_{\beta} F_{n,i}) \cdot \|F_{n,j}\|_1 \\ &\leq C \sup_{1 \leq j \leq k_n} \left(\rho_n k_n^2 \frac{\|F_{n,j}\|_1^2}{\check{s}_n^2} + \frac{\rho_n}{1 - r_n} \sum_{1 \leq i \leq k_n} r^{n_i} \frac{D_{\beta} F_{n,i}}{\check{s}_n} \frac{\|F_{n,j}\|_1}{\check{s}_n} \right) \\ &\leq C \sup_{1 \leq j \leq k_n} \left(\rho_n k_n^2 \frac{\|F_{n,j}\|_2^2}{\check{s}_n^2} + \frac{\rho_n}{1 - r_n} \sum_{1 \leq i \leq k_n} r^{n_i} \frac{\|F_{n,i}\|}{\check{s}_n} \frac{\|F_{n,j}\|_2}{\check{s}_n} \right) \\ &\stackrel{(18)}{\leq} C \rho_n k_n^2 (\epsilon^2 + L_{n,\epsilon}) + \frac{\rho_n}{1 - r_n} \sum_{1 \leq i \leq k_n} r^{n_i} \frac{\|F_{n,i}\|}{\check{s}_n} (\epsilon^2 + L_{n,\epsilon})^{1/2}. \end{aligned}$$

Now the assumptions (11), (12) and (13) imply that the limsup of the upper bound is bounded by $K\epsilon$ for some $K > 0$, hence (17) follows. \square

Proof of Theorem 5.3. Extending our probability space if necessary, we may assume that there exists an array of random variables $\{X_{n,i}\}_{i=1}^{k_n}$ such that $X_{n,i}, i = 1, \dots, k_n$, are independent normal random variables and

$$\mathbb{E}X_{n,i} = 0 \text{ and } \text{Var } X_{n,i} = \text{Var } F_{n,i}. \quad (19)$$

Without loss of generality we may as well assume that for each n , $\{X_{n,i}\}_{i=1}^{k_n}$ and $\{F_{n,i} \circ T^{\tau_{n,i}}\}_{i=1}^{k_n}$ are independent. Define two random variables

$$F_n = \frac{F_{n,1} \circ T^{\tau_{n,1}} + F_{n,2} \circ T^{\tau_{n,2}} + \cdots + F_{n,k_n} \circ T^{\tau_{n,k_n}}}{\check{s}_n},$$

$$X_n = \frac{X_{n,1} + \cdots + X_{n,k_n}}{\check{s}_n}.$$

X_n is a normal random variable for being a sum of normal random variables and converges weakly to $\mathcal{N}(0, 1)$ because of (17). Since F_n has variance 1, the set of distributions of F_n is mass-preserving. To show that F_n also converges weakly to $\mathcal{N}(0, 1)$, it suffices to prove that for any h in the separating class $C_c^\infty(\mathbb{R})$,

$$\mathbb{E}h(F_n) - \mathbb{E}h(X_n) \rightarrow 0.$$

Letting

$$U_{n,i} := \frac{F_{n,1} \circ T^{\tau_{n,1}} + \cdots + F_{n,i-1} \circ T^{\tau_{n,i-1}}}{\check{s}_n} + \frac{X_{n,i+1} + \cdots + X_{n,k_n}}{\check{s}_n},$$

we can write

$$h(F_n) - h(X_n) = \sum_{i=1}^{k_n} h\left(U_{n,i} + \frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n}\right) - h\left(U_{n,i} + \frac{X_{n,i}}{\check{s}_n}\right).$$

Use Taylor expansion to deduce that

$$h(F_n) - h(X_n) = \sum_{i=1}^{k_n} h'(U_{n,i}) \left(\frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n} - \frac{X_{n,i}}{\check{s}_n} \right) + h''\left(U_{n,i} + \theta_{n,i} \frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n}\right) \frac{F_{n,i}^2 \circ T^{\tau_{n,i}}}{2\check{s}_n^2} - h''\left(U_{n,i} + \tilde{\theta}_{n,i} \frac{X_{n,i}}{\check{s}_n}\right) \frac{X_{n,i}^2}{2\check{s}_n^2},$$

where $\theta_{n,i}, \tilde{\theta}_{n,i} : \Omega \rightarrow [0, 1]$. Rewrite the right-hand side as

$$\begin{aligned} & \sum_{i=1}^{k_n} h'(U_{n,i}) \left(\frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n} - \frac{X_{n,i}}{\check{s}_n} \right) + h''(U_{n,i}) \left(\frac{F_{n,i}^2 \circ T^{\tau_{n,i}}}{2\check{s}_n^2} - \frac{X_{n,i}^2}{2\check{s}_n^2} \right) \\ & + \left\{ h''\left(U_{n,i} + \theta_{n,i} \frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n}\right) \frac{F_{n,i}^2 \circ T^{\tau_{n,i}}}{2\check{s}_n^2} - h''(U_{n,i}) \frac{F_{n,i}^2 \circ T^{\tau_{n,i}}}{2\check{s}_n^2} \right\} \\ & - \left\{ h''\left(U_{n,i} + \tilde{\theta}_{n,i} \frac{X_{n,i}}{\check{s}_n}\right) \frac{X_{n,i}^2}{2\check{s}_n^2} - h''(U_{n,i}) \frac{X_{n,i}^2}{2\check{s}_n^2} \right\}. \quad (20) \end{aligned}$$

We are about to show that the expectation of (20) vanishes asymptotically. Denote by

$$\mathbb{E}_{n,i}(\cdot) := \mathbb{E}(\cdot | F_{n,1} \circ T^{\tau_{n,1}}, \dots, F_{n,i} \circ T^{\tau_{n,i}})$$

the corresponding conditional expectation. To estimate the expectation of the first summand in (20), we write

$$\begin{aligned}
& \mathbb{E} \left(\sum_{i=1}^{k_n} h'(U_{n,i}) \left(\frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n} - \frac{X_{n,i}}{\check{s}_n} \right) \right) \\
&= \sum_{i=1}^{k_n} \mathbb{E} \left(\mathbb{E}_{n,i} (h'(U_{n,i})) \cdot \frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n} \right) - \mathbb{E} h'(U_{n,i}) \mathbb{E} \frac{X_{n,i}}{\check{s}_n} \\
&\stackrel{(19)}{=} \frac{1}{\check{s}_n} \sum_{i=2}^{k_n} \int_{\Omega} \mathcal{L}^{\tau_{n,i}} \mathbb{E}_{n,i} h'(U_{n,i}) \cdot F_{n,i} d\mu \\
&\stackrel{(10)}{=} \frac{1}{\check{s}_n} \sum_{i=2}^{k_n} \left(\mathbb{E} h'(U_{n,i}) \int_{\Omega} F_{n,i} d\mu + \int_{\Omega} N^{\tau_{n,i}} \mathbb{E}_{n,i} h'(U_{n,i}) \cdot F_{n,i} d\mu \right) \\
&= \frac{1}{\check{s}_n} \sum_{i=2}^{k_n} \int_{\Omega} N^{\tau_{n,i}} \mathbb{E}_{n,i} h'(U_{n,i}) \cdot F_{n,i} d\mu, \tag{21}
\end{aligned}$$

where in the first equality we use the independence between $U_{n,i}$ and $X_{n,i}$ and

$$\mathbb{E}_{n,i} (h'(U_{n,i}) \cdot F_{n,i} \circ T^{\tau_{n,i}}) = \mathbb{E}_{n,i} (h'(U_{n,i})) \cdot F_{n,i} \circ T^{\tau_{n,i}},$$

and the last equality is due to $\int_{\Omega} F_{n,i} d\mu = 0$. Observe the following inequalities.

1. By (15),

$$\begin{aligned}
& \frac{1}{\check{s}_n} \sum_{i=2}^{k_n} \left| \int_{\Omega} N^{\tau_{n,i}} \mathbb{E}_{n,i} h'(U_{n,i}) \cdot F_{n,i} d\mu \right| \\
&\leq C \sum_{i=2}^{k_n} (\rho_n \|\mathbb{E}_{n,i} h'(U_{n,i})\|_1 + \rho_n r^{\tau_{n,i-1} + n_{i-1}} D_{\beta}(\mathbb{E}_{n,i} h'(U_{n,i}))) \cdot \frac{\|F_{n,i}\|_1}{\check{s}_n}.
\end{aligned}$$

2.

$$\|\mathbb{E}_{n,i} h'(U_{n,i})\|_1 \leq \mathbb{E}(\mathbb{E}_{n,i} |h'(U_{n,i})|) \leq \|h'\|_{\infty}.$$

3. Recall that $U_{n,i} = \frac{1}{\check{s}_n} \left(\sum_{j=1}^{i-1} F_{n,j} \circ T^{\tau_{n,j}} + \sum_{j=i+1}^{k_n} X_{n,j} \right)$. Because $\{X_{n,j}\}_{j=1}^{k_n}$ and $\{F_{n,j} \circ T^{\tau_{n,j}}\}_{j=1}^{k_n}$ are independent,

$$D_{\beta}(\mathbb{E}_{n,i} h'(U_{n,i})) \leq \frac{\|h''\|_{\infty}}{\check{s}_n} D_{\beta}(F_{n,1} \circ T^{\tau_{n,1}} + \dots + F_{n,i-1} \circ T^{\tau_{n,i-1}}).$$

4. For any $f \in L$ and $m \in \mathbb{N}$,

$$\begin{aligned}
D_\beta(f \circ T^m) &= \sup_{b \in \beta, x, y \in b} \frac{|f \circ T^m(x) - f \circ T^m(y)|}{r(x, y)} \\
&= \sup_{b \in \beta, x, y \in b} \frac{|f \circ T^m(x) - f \circ T^m(y)|}{r(T^m x, T^m y)} \frac{r(T^m x, T^m y)}{r(x, y)} \\
&\leq D_\Omega f \cdot r^{-m} \\
&\leq \max\left\{\frac{2\|f\|_\infty}{r}, D_\beta(f)\right\} \cdot r^{-m} = O(1)\|f\| r^{-m}. \quad (22)
\end{aligned}$$

We use these inequalities to estimate (21),

$$\begin{aligned}
&\frac{1}{\check{s}_n} \sum_{i=2}^{k_n} \left| \int_{\Omega_n} N^{\tau_{n,i}} \mathbb{E}_{n,i} h'(U_{n,i}) \cdot F_{n,i} d\mu \right| \\
&\leq O(1) \sum_{i=2}^{k_n} \left(\rho_n \|h'\|_\infty + \rho_n r^{\tau_{n,i-1} + n_{i-1}} \frac{\|h''\|_\infty}{\check{s}_n} \sum_{j=1}^{i-1} \frac{1}{r^{\tau_{n,j}}} \|F_{n,j}\| \right) \cdot \frac{\|F_{n,i}\|_1}{\check{s}_n} \\
&\leq O(1) \left(k_n \rho_n + \rho_n \sum_{i=2}^{k_n} \sum_{j=1}^{i-1} r_n^{i-j-1} r^{n_j + \dots + n_{i-1}} \frac{\|F_{n,j}\|}{\check{s}_n} \right) \cdot \sup_{1 \leq i \leq k_n} \frac{\|F_{n,i}\|_2}{\check{s}_n} \\
&\leq O(1) \left(k_n \rho_n + \frac{\rho_n}{1 - r_n} \sum_{1 \leq j \leq k_n} r^{n_j} \frac{\|F_{n,j}\|}{\check{s}_n} \right) \cdot \sup_{1 \leq i \leq k_n} \frac{\|F_{n,i}\|_2}{\check{s}_n}. \quad (23)
\end{aligned}$$

The bound tends to 0 as $n \rightarrow \infty$ because of (16) and assumptions (11) and (12).

The expectation of the second summand in (20) is estimated in a similar way. We rewrite

$$\begin{aligned}
&\mathbb{E} \sum_{i=1}^{k_n} h''(U_{n,i}) (F_{n,i}^2 \circ T^{\tau_{n,i}} - X_{n,i}^2) \\
&= \sum_{i=2}^{k_n} \int_{\Omega} \mathbb{E}_{n,i} h''(U_{n,i}) \cdot F_{n,i}^2 \circ T^{\tau_{n,i}} d\mu - \mathbb{E} h''(U_{n,i}) \text{Var } X_{n,i} \\
&\stackrel{(10)}{=} \sum_{i=2}^{k_n} \mathbb{E} h''(U_{n,i}) \text{Var } F_{n,i} + \int_{\Omega} N^{\tau_{n,i}} \mathbb{E}_{n,i} h''(U_{n,i}) \cdot F_{n,i}^2 d\mu - \mathbb{E} h''(U_{n,i}) \text{Var } X_{n,i} \\
&\stackrel{(19)}{=} \sum_{i=2}^{k_n} \int_{\Omega} N^{\tau_{n,i}} \mathbb{E}_{n,i} h''(U_{n,i}) \cdot F_{n,i}^2 d\mu.
\end{aligned}$$

Then we can repeat the estimate for (21) to deduce an upper-bound similar to (23).

The expectation of the third summand in (20) is equal to

$$\begin{aligned}
& \sum_{i=1}^{k_n} \int_{\Omega} \left(h'' \left(U_{n,i} + \theta_{n,i} \frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n} \right) - h''(U_{n,i}) \right) \frac{F_{n,i}^2 \circ T^{\tau_{n,i}}}{2\check{s}_n^2} d\mu \\
&= \sum_{i=1}^{k_n} \int_{\Omega} \left(h'' \left(U_{n,i} + \theta_{n,i} \frac{F_{n,i} \circ T^{\tau_{n,i}}}{\check{s}_n} \right) - h''(U_{n,i}) \right) \frac{F_{n,i}^2 \circ T^{\tau_{n,i}}}{2\check{s}_n^2} \\
&\quad \left(\mathbf{1}_{\{|F_{n,i} \circ T^{\tau_{n,i}}| < \epsilon \check{s}_n\}} + \mathbf{1}_{\{|F_{n,i} \circ T^{\tau_{n,i}}| \geq \epsilon \check{s}_n\}} \right) d\mu \\
&\leq \epsilon \|h'''\|_{\infty} \sum_{i=1}^{k_n} \int_{\Omega} \frac{F_{n,i}^2}{\check{s}_n^2} d\mu + \|h''\|_{\infty} L_{n,\epsilon}
\end{aligned}$$

for any $\epsilon > 0$. This expectation converges to 0 in view of (17) and (13). The expectation of the last summand in (20) is controlled in the same way as the third summand. \square

Applying this theorem to Birkhoff sums, we obtain the following result.

Corollary 5.6. *Given a Gibbs-Markov system (Ω, μ, T, α) and a sequence of centered functions $\{f_n\}$ in L . Let $s_n^2 := \text{Var}(f_n + \dots + f_n \circ T^{n-1})$. Assume that there are sequences of integers $\tau_n > m_n$ with the following extra properties.*

1.

$$\limsup_{n \rightarrow \infty} k_n^2 \rho^{m_n} < \infty,$$

$$\text{where } k_n := \left\lfloor \frac{n}{\tau_n + m_n} \right\rfloor.$$

2.

$$\limsup_{n \rightarrow \infty} k_n \rho^{m_n} \frac{\|f_n\|}{s_n} < \infty. \quad (24)$$

3. For every $1 \leq i \leq \tau_n$

$$\frac{1}{s_n^2} \int_{\Omega} (f_n + \dots + f_n \circ T^{i-1})^2 d\mu \rightarrow 0. \quad (25)$$

4.

$$\frac{k_n}{s_n^2} \int_{\Omega} (f_n + \dots + f_n \circ T^{m_n-1})^2 d\mu \rightarrow 0. \quad (26)$$

5.

$$\frac{k_n}{s_n^2} (f_n + \dots + f_n \circ T^{\tau_n-1})^2 \text{ is uniformly integrable.} \quad (27)$$

Then $\frac{f_n + \dots + f_n \circ T^{n-1}}{s_n} \Rightarrow \mathcal{N}(0, 1)$.

Proof. Let $F_{n,i} := f_n + \dots + f_n \circ T^{\tau_n-1}$, $g_{n,i} := f_n + \dots + f_n \circ T^{m_n-1}$ and $\tau_{n,i} := (i-1)(\tau_n + m_n)$ for $1 \leq i \leq k_n$, then

$$f_n + \dots + f_n \circ T^{k_n(\tau_n+m_n)-1} = \sum_{i=1}^{k_n} F_{n,i} \circ T^{\tau_{n,i}} + g_{n,i} \circ T^{\tau_{n,i}+\tau_n}.$$

To complete the ergodic sum, let $F_{n,k_n+1} := f_n + \dots + f_n \circ T^{\min\{\tau_n, n-k_n(\tau_n+m_n)\}-1}$, $\tau_{n,k_n+1} := k_n(\tau_n + m_n)$ and $g_{n,k_n+1} := f_n + \dots + f_n \circ T^{n-k_n(\tau_n+m_n)-\tau_n-1}$ if necessary. The following two properties ensure that the dynamical array $\{(F_{n,i}, \tau_{n,i}) : i = 1, \dots, k_n + 1\}$ has the same distributional limit as the ergodic sum.

$$1. \quad \frac{\check{s}_n}{s_n} \rightarrow 1, \quad (28)$$

where $\check{s}_n^2 = \text{Var} \sum_{i=1}^{k_n+1} F_{n,i} \circ T^{\tau_{n,i}}$.

$$2. \quad \frac{\sum_{i=1}^{k_n+1} g_{n,i} \circ T^{\tau_{n,i}+\tau_n}}{\check{s}_n} \Rightarrow 0. \quad (29)$$

In fact, to see (28) first note that

$$s_n^2 = \check{s}_n^2 + \text{Var} \sum_{i=1}^{k_n+1} g_{n,i} \circ T^{\tau_{n,i}+\tau_n} + 2 \int_{\Omega} \sum_{i=1}^{k_n+1} F_{n,i} \circ T^{\tau_{n,i}} \cdot \sum_{i=1}^{k_n+1} g_{n,i} \circ T^{\tau_{n,i}+\tau_n} d\mu.$$

With conditions (24) and (26), arguments involving the transfer operator similar to those used in proving (17) indicate that

$$\lim_{n \rightarrow \infty} \frac{\text{Var} \sum_{i=1}^{k_n+1} g_{n,i} \circ T^{\tau_{n,i}+\tau_n}}{s_n^2} = \lim_{n \rightarrow \infty} \sum_{i=1}^{k_n+1} \frac{\int_{\Omega} g_{n,i}^2 d\mu}{s_n^2},$$

which is 0 by (25) and (26). Since

$$\begin{aligned} \int_{\Omega} F_{n,i} \circ T^{\tau_{n,i}} \cdot \sum_{j=1}^{k_n+1} g_{n,j} \circ T^{\tau_{n,j}+\tau_n} d\mu &= \sum_{j>i} \int_{\Omega} F_{n,i} \cdot g_{n,j} \circ T^{\tau_{n,j}+\tau_n-\tau_{n,i}} d\mu + \\ \sum_{j<i-1} \int_{\Omega} g_{n,j} \cdot F_{n,i} \circ T^{\tau_{n,i}-\tau_{n,j}-\tau_n} d\mu &+ \int_{\Omega} F_{n,i} \cdot g_{n,i} \circ T^{\tau_n} d\mu + \int_{\Omega} g_{n,i-1} \cdot F_{n,i} \circ T^{m_n} d\mu, \end{aligned}$$

one can apply the same arguments using the transfer operator repeatedly to show that

$$\frac{1}{s_n^2} \int_{\Omega} \sum_{i=1}^{k_n+1} F_{n,i} \circ T^{\tau_{n,i}} \cdot \sum_{j=1}^{k_n+1} g_{n,j} \circ T^{\tau_{n,j}+\tau_n} d\mu \rightarrow 0.$$

Thus (28) holds. The previous arguments also imply that (29) is just a consequence of (26). Hence we only need to verify the conditions in Theorem 5.3 for the dynamical array $\{(F_{n,i}, \tau_{n,i}) : i = 1, \dots, k_n + 1\}$. (12) is taken care of by (24) since

$$\limsup_{n \rightarrow \infty} \rho^{m_n} \sum_{1 \leq i \leq k_n} r^{\tau_n} \frac{\|F_{n,i}\|}{\check{s}_n} \leq \limsup_{n \rightarrow \infty} \rho^{m_n} k_n \frac{\|f_n\|}{\check{s}_n}.$$

Note that

$$\sum_{i=1}^{k_n+1} \int_{\Omega} \frac{F_{n,i}^2}{\check{s}_n^2} \mathbf{1}_{\{|F_{n,i}| \geq \epsilon \check{s}_n\}} d\mu = \int_{\Omega} k_n \frac{F_{n,1}^2}{\check{s}_n^2} \mathbf{1}_{\{|F_{n,1}| \geq \epsilon \check{s}_n\}} + \frac{F_{n,k_n+1}^2}{\check{s}_n^2} \mathbf{1}_{\{|F_{n,k_n+1}| \geq \epsilon \check{s}_n\}} d\mu$$

hence the Lindeberg condition (13) follows from (27), (25) and (28). \square

Remark 5.7. Theorem 5.3 also can be generalized, with the same assumptions, to more general dynamical systems. It in fact holds for any system for which the transfer operator satisfies the Doeblin-Fortet inequality (4) and for which the composition operator satisfies the inequality (22) (or in the case of Lipschitz norm, $r(T^m x, T^m y) \leq \frac{r(x,y)}{r^m}$). Note that the inequalities (4) and (22) are bounded at the rates of r and r^{-1} respectively.

6 Applications to the large sample theory in statistics

The central limit theorem under the Lindeberg condition has many applications, in particular in nonparametric statistics. The book [14] provides a glimpse on these applications, though it is not a complete list. Here we restrict to one particular case, the famous Behrens-Fisher problem. More applications will be described elsewhere.

We will use the setup for the two sample problem in a Gibbs-Markov dynamical system (Ω, μ, T, α) . Let $\phi, \psi : \Omega \rightarrow \mathbb{R}$ be two functions in the class L , which determine two stationary sequences $X_n = \phi \circ T^n$ and $Y_n = \psi \circ T^n$. For simplicity we assume that the distributions μ_ϕ of ϕ and μ_ψ of ψ have no atoms. Based on observations X_1, \dots, X_m and Y_1, \dots, Y_n , the Behrens-Fisher problem is to determine whether the distributions of ϕ and ψ are different or not in a statistical sense. We shall deal with this problem when the distributions differ in their means, that is $\int \phi d\mu \neq \int \psi d\mu$.

The classical solution for this problem (to be the most powerful test) is the t -test which only works exactly under normal distribution, independence and equal variances. In all other cases some type of approximation is needed. In particular, when the distributions of ϕ and ψ are completely unknown, the two sample Wilcoxon rank sum test is widely used. Consider $m, n \in \mathbb{N}$ and observations X_1, \dots, X_m and Y_1, \dots, Y_n . Define R_i to be the rank of X_i among

all $n + m$ observations $X_1, \dots, X_m, Y_1, \dots, Y_n$. Then

$$W_{m,n} = \sum_{i=1}^m R_i$$

is the two sample Wilcoxon rank sum test. In order to solve the problem in a nonparametric setup one needs to determine the asymptotic distribution of $W_{m,n}$.

$$\begin{aligned} W_{m,n} &= \sum_{i=1}^m \sum_{k=1}^n \mathbf{1}_{\{Y_k \leq X_i\}} + \sum_{i=1}^m \sum_{k=1}^m \mathbf{1}_{\{X_k \leq X_i\}} \\ &= \sum_{i=1}^m \sum_{k=1}^n \mathbf{1}_{\{Y_k \leq X_i\}} + \frac{m(m-1)}{2} \\ &= \left(\sum_{i=1}^m \sum_{k=1}^n \left(\mathbf{1}_{\{Y_k \leq X_i\}} - \int \mathbf{1}_{\{Y_k \leq t\}} \mu_\phi(dt) - \int \mathbf{1}_{\{t \leq X_i\}} \mu_\psi(dt) + \iint \mathbf{1}_{\{s \leq t\}} \mu_\phi(dt) \mu_\psi(ds) \right) \right) \\ &\quad + m \left(\sum_{k=1}^n \int \mathbf{1}_{\{Y_k \leq t\}} \mu_\phi(dt) - n \iint \mathbf{1}_{\{s \leq t\}} \mu_\phi(dt) \mu_\psi(ds) \right) \\ &\quad + n \left(\sum_{i=1}^m \int \mathbf{1}_{\{t \leq X_i\}} \mu_\psi(dt) - m \iint \mathbf{1}_{\{s \leq t\}} \mu_\phi(dt) \mu_\psi(ds) \right) \\ &\quad + \left(mn \iint \mathbf{1}_{\{s \leq t\}} \mu_\phi(dt) \mu_\psi(ds) + \frac{m(m-1)}{2} \right) \\ &=: A + mB + nC + D. \end{aligned}$$

We first give conditions under which the second moment of A , normalized by m^3 converges to zero as $m \rightarrow \infty$ and $n/m \rightarrow \lambda \in (0, 1)$. This can be seen directly or by applying [9] when $(x, y) \mapsto \mathbf{1}_{\{\psi(y) \leq \phi(x)\}}$ approximately belongs to the projective tensor product $L_{2,\pi}(\mu^2)$ over $L_2(\mu^2)$ and therefore the variance of the approximation \tilde{A} of A increases like $\sqrt{nm} \|\tilde{A}\|_{L_{2,\pi}(\mu^2)}$. We refer to [9] for the definitions and properties of projective tensor products. Alternatively, assuming that the distributions of ψ and ϕ are absolutely continuous with respect to Lebesgue measure and have a bounded density, one could use Theorem 1 (actually Lemma 3) in [10] to show that the variance of A is of smaller order. We prove

Proposition 6.1. *Assume that the distributions μ_ϕ and μ_ψ satisfy*

$$\mu_\phi(I) \leq K\eta^r \quad \text{and} \quad \mu_\psi(I) \leq K\eta^r \quad \forall \eta > 0, \text{ interval } I \text{ of length } \eta$$

for some $K > 0$ and $r \in (\frac{4}{5}, 4)$ and assume that $\|\phi\|_\infty$ and $\|\psi\|_\infty$ are finite. Then, as $n/m \rightarrow \lambda \in (0, 1)$, A has a representation $A = A_1 + A_2$ so that

$$\text{Var } A_1 = o(m^3) \quad \text{and} \quad \mathbb{E}|A_2| = o(m^{3/2}).$$

Therefore, normalized by $m^{-3/2}$, A does not contribute to the distributional limit of $W_{m,n}$.

Proof. Let $m \in \mathbb{N}$ and choose q which depends on m and is chosen below. Let $M = \max\{\|\phi\|_\infty, \|\psi\|_\infty\}$ and $h(x, y) = \mathbf{1}_{\{-M \leq y \leq x \leq M\}}$. Divide the interval $[-M, M]$ into q subintervals J_1, \dots, J_q of equal length $2Mq^{-1}$ and let

$$I_j = \{(x, y) : x \in J_j, -M \leq y \leq \min J_j\}, \quad I = \bigcup_{j=1}^q I_j.$$

Then $\mu_\phi \times \mu_\psi(\{(x, y) : -M \leq y \leq x \leq M\} \setminus I) \leq K(2M)^r q^{-r}$ and the projective norm of $(u, v) \mapsto \tilde{h}_q(u, v) = \mathbf{1}_I(\phi(u), \psi(v))$ is bounded by (cf. [9, Lemma 1])

$$\|\tilde{h}_q\|_{L_{4,\pi}(\mu^2)} \leq \sum_{j=1}^q \|\mathbf{1}_{I_j}(\phi, \psi)\|_{L_4(\mu^2)} \leq q [K(2M)^r q^{-r}]^{1/4}.$$

Write

$$\hat{h}_q(u, v) = \tilde{h}_q(u, v) - \int \tilde{h}_q(w, v) \mu(dw) - \int \tilde{h}_q(u, w) \mu(dw) + \iint \tilde{h}_q(w, w') \mu(dw) \mu(dw')$$

and

$$A_1 = \sum_{k=1}^n \sum_{i=1}^m \hat{h}_q(T^i(u), T^k(v)).$$

Then $\|\hat{h}_q\|_{L_{4,\pi}(\mu^2)} = O(\|\tilde{h}_q\|_{L_{4,\pi}(\mu^2)})$ and applying Lemma 4 in [9] with $d = m = 2$ and $p = 4$ (one can verify the assumption in this lemma for \hat{h}_q) it follows that there is a constant C (independent of q and (n, m)) such that

$$\|A_1\|_{L_2(\mu)} \leq C\sqrt{nm} \|\hat{h}_q\|_{L_{4,\pi}(\mu^2)} = O(mq^{1-r/4}).$$

Moreover, we get

$$\iint \left| \sum_{k=1}^n \sum_{i=1}^m h(\phi(T^i(u)), \psi(T^k(v))) - \tilde{h}_q(T^i(u), T^k(v)) \right| \mu(du) \mu(dv) \leq K(2M)^r q^{-r} nm.$$

Similar estimates hold for the other summand in $A_2 = A - A_1$.

Since $4 > r > \frac{4}{5}$ we have that $0 < 2 - \frac{r}{2} < 2r$, hence can pick

$$\frac{1}{2r} < \tau < \frac{1}{2 - \frac{r}{2}}$$

and $q = m^\tau$ to obtain

$$m^{-3} \mathbb{E}(A_1^2) = O(m^{-3} m^2 q^{2-r/2}) = O(m^{-1+2\tau-\frac{r\tau}{2}}) = o(1)$$

and

$$m^{-3/2} \mathbb{E}|A - A_1| = O(m^{-3/2} m^2 q^{-r}) = O(m^{\frac{1}{2}-r\tau}) = o(1).$$

□

The variances of the terms B and C grow approximately linearly with n (resp. m) as can be seen e.g. from Lemma 3.6, when the distribution functions of μ_ϕ and μ_ψ are assumed to belong to the class L .

We are not developing more details and extensions of the forgoing discussions. Instead, in what follows, let us assume that

$$\text{Var}(A) = o(m^3),$$

and that the distributions of ψ and ϕ are absolutely continuous with bounded densities. Under these simplifying assumptions the following argument becomes short and shows the general pattern of proof under more general assumptions.

Proposition 6.2. *Under the assumptions explained above*

$$m^{-3/2} (W_{m,n} - D) \Rightarrow \mathcal{N}(0, \sigma^2)$$

weakly for some $\sigma^2 \geq 0$ as $m \rightarrow \infty$ and $n/m \rightarrow \lambda$ for some $0 < \lambda < 1$.

Note that in case the distributions of ϕ and ψ are equal, then

$$D = \frac{mn}{2} + \frac{m(m-1)}{2} = \frac{m}{2}(n+m-1).$$

This shows that the two sample Wilcoxon rank sum test checks whether the distributions of ϕ and ψ differ by a location alternative.

Proof. The assertion of the proposition follows if for any $s, t \in \mathbb{R}$

$$V(s, t) = \frac{s}{\sqrt{n}}B + \frac{t}{\sqrt{m}}C$$

converges weakly to some normal distribution with expectation 0 and some variance $\sigma^2(s, t) \geq 0$ (Cramér-Wold device) which is the variance of $sU + tV$ for some two-dimensional normal variable (U, V) .

This can be shown using the Lindeberg central limit theorem 5.3. Let $F_\psi(x) = \mu_\psi((-\infty, x])$ and $F_\phi(x) = \mu_\phi((-\infty, x])$. Then,

$$\begin{aligned} \frac{s}{\sqrt{n}}B + \frac{t}{\sqrt{m}}C &= \frac{s}{\sqrt{n}} \sum_{k=1}^n \left[1 - F_\phi(\psi \circ T^k) - \int (1 - F_\phi(x)) F_\psi(dx) \right] \\ &\quad + \frac{t}{\sqrt{m}} \sum_{i=1}^m \left[F_\psi(\phi \circ T^i) - \int F_\psi(x) F_\phi(dx) \right]. \end{aligned}$$

Since F_ψ and F_ϕ are Lipschitz continuous with Lipschitz constants bounded by the supremum of their density functions, the variance of $V(s, t)$ is asymptotic to $\sigma^2(s, t)$ with

$$\begin{aligned} \sigma^2(s, t) &= \lim_{m \rightarrow \infty, n/m \rightarrow \lambda} \frac{s^2}{n} \text{Var} \left(\sum_{k=0}^{n-1} F_\phi(\psi \circ T^k) \right) \\ &\quad - \frac{st}{\sqrt{nm}} \text{Cov} \left(\sum_{k=0}^{n-1} F_\phi(\psi \circ T^k), \sum_{i=0}^{m-1} F_\psi(\phi \circ T^i) \right) + \frac{t^2}{m} \text{Var} \left(\sum_{i=0}^{m-1} F_\psi(\phi \circ T^i) \right). \end{aligned} \tag{30}$$

We can apply Theorem 5.3 setting $\theta = \iint \mathbf{1}_{\{\psi \leq \phi\}} d\mu_\psi d\mu_\phi$, $n = k_n(p+q) + q_n$ and $m = k_m(p+q) + q_m$ where $0 \leq q_n, q_m < p+q$. Denote

$$\begin{aligned} F_{n,i} &= \sum_{l=0}^{p-1} \left[\frac{t}{\sqrt{m}} F_\psi(\phi \circ T^l) - \frac{s}{\sqrt{n}} F_\phi(\psi \circ T^l) - \left(\frac{t}{\sqrt{m}} - \frac{s}{\sqrt{n}} \right) \theta \right] \quad 1 \leq i \leq k_n, \\ F_{n,i} &= \sum_{l=0}^{p-1} \left[\frac{t}{\sqrt{m}} F_\psi(\phi \circ T^l) - \frac{t}{\sqrt{m}} \theta \right] \quad i = k_n + 1, \dots, k_m, \\ \tau_{n,i} &= (i-1)(p+q) \quad i = 1, \dots, k_m. \end{aligned}$$

We check next that conditions 1.-4. in Theorem 5.3 hold with an appropriate choice of p and q . If $\sigma^2(s, t) = 0$ we have convergence to 0 in probability, hence nothing has to be shown. Assume that this is not the case, which ensures that 1. is satisfied. Let r and ρ be given by the transformation T . Choosing $p = O(m^{\frac{1}{2}-\delta})$ for some $\frac{1}{6} < \delta < \frac{1}{2}$ and $q = o(\frac{2 \log m}{-\log \rho})$ it follows that

$$k_m^2 \rho^q = o(1),$$

hence 2. holds. Since $\|F_{n,i}\| \leq r^{-p}(\frac{t}{\sqrt{m}} D_\alpha F_\psi + \frac{s}{\sqrt{n}} D_\alpha F_\phi) + \sqrt{\text{Var}(F_{n,i})}$ for $i = 1, \dots, k_n$ and similarly for $i = k_n + 1, \dots, k_m$ condition 3. is satisfied:

$$\rho^q k_m r^p \|F_{n,1}\| \check{s}_n^{-1} = o(m^{-\frac{3}{2}+\delta} r^p (r^{-p} m^{-1/2} + p m^{-1/2})) = o(1).$$

Finally, the Lindeberg condition 4. holds since

$$\int F_{n,i}^2 \mathbf{1}_{\{|F_{n,i}| \geq \epsilon \check{s}_n\}} d\mu \leq \frac{1}{\epsilon^2 \check{s}_n^2} \int F_{n,i}^4 d\mu \leq O(m^{-1} p^2) \frac{1}{\epsilon^2 \check{s}_n^2} \int F_{n,i}^2 d\mu$$

and summing over i and dividing by \check{s}_n^2 yields a bound of $(m^{-1} p^2)^2 k_m \check{s}_n^{-4} = O(m^{\frac{1}{2}-3\delta})$.

It is proved now that

$$\frac{1}{\check{s}_n} \sum_{i=1}^{k_m} F_{n,i}$$

converges to the distribution of $sU + tV$ where (U, V) is normal with expectation 0 and covariance given by the single summands in (30) for $s = t = 1$. It is therefore left to show that $\frac{1}{\check{s}_n} \sum_{i=1}^{k_m} F_{n,i}$ is stochastically equivalent to $\frac{1}{\check{s}_n} V(s, t)$. The variance of the difference is bounded by

$$\check{s}_n^{-2} ((k_m + 2) q^2 \left(\frac{s}{\sqrt{n}} + \frac{t}{\sqrt{m}} \right)^2) = O(m^{\frac{1}{2}+\delta} (\log m)^2 m^{-1}) = o(1)$$

where we use the fact that its variance grows linearly in k_m , the number of block-sums forming the difference. This finishes the proof. \square

Acknowledgment: The research was supported by *número 158/2012 de Pesquisador Visitante Especial de CAPES*.

References

- [1] Jon Aaronson and Manfred Denker. The Poincaré series of $\mathbf{C} \setminus \mathbf{Z}$. *Ergodic Theory Dynam. Systems*, 19(1):1–20, 1999.
- [2] Jon Aaronson and Manfred Denker. Local limit theorems for partial sums of stationary sequences generated by Gibbs-Markov maps. *Stoch. Dyn.*, 1(2):193–237, 2001.
- [3] Jon Aaronson, Manfred Denker, and Mariusz Urbański. Ergodic theory for Markov fibred systems and parabolic rational maps. *Trans. Amer. Math. Soc.*, 337(2):495–548, 1993.
- [4] Jean-René Chazottes. Fluctuations of observables in dynamical systems: from limit theorems to concentration inequalities. In *Nonlinear dynamics new directions*, volume 11 of *Nonlinear Syst. Complex.*, pages 47–85. Springer, Cham, 2015.
- [5] Guy Cohen and Jean-Pierre Conze. CLT for random walks of commuting endomorphisms on compact abelian groups. *to appear in Journal of Theoretical Probability*.
- [6] Jean-Pierre Conze and Albert Raugi. Limit theorems for sequential expanding dynamical systems on $[0, 1]$. In *Ergodic theory and related fields*, volume 430 of *Contemp. Math.*, pages 89–121. Amer. Math. Soc., Providence, RI, 2007.
- [7] Manfred Denker. Statistical decision procedures and ergodic theory. In *Ergodic theory and related topics (Vitte, 1981)*, volume 12 of *Math. Res.*, pages 35–47. Akademie-Verlag, Berlin, 1982.
- [8] Manfred Denker. The central limit theorem for dynamical systems. In *Dynamical systems and ergodic theory (Warsaw, 1986)*, volume 23 of *Banach Center Publ.*, pages 33–62. PWN, Warsaw, 1989.
- [9] Manfred Denker and Mikhail Gordin. Limit theorems for von Mises statistics of a measure preserving transformation. *Probab. Theory Related Fields*, 160(1-2):1–45, 2014.
- [10] Manfred Denker and Gerhard Keller. Rigorous statistical procedures for data from dynamical systems. *J. Statist. Phys.*, 44(1-2):67–93, 1986.
- [11] Manfred Denker, Feliks Przytycki, and Mariusz Urbański. On the transfer operator for rational functions on the Riemann sphere. *Ergodic Theory Dynam. Systems*, 16(2):255–266, 1996.
- [12] Manfred Denker, Samuel Senti, and Xuan Zhang. Fluctuations of ergodic sums over periodic orbits under specification. preprint, 2015.

- [13] Nelson Dunford and Jacob T. Schwartz. *Linear Operators. I. General Theory*. With the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, Vol. 7. Interscience Publishers, Inc., New York; Interscience Publishers, Ltd., London, 1958.
- [14] Thomas S. Ferguson. *A course in large sample theory*. Texts in Statistical Science Series. Chapman & Hall, London, 1996.
- [15] M. I. Gordin. The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR*, 188:739–741, 1969.
- [16] Sébastien Gouëzel. Characterization of weak convergence of Birkhoff sums for Gibbs-Markov maps. *Israel J. Math.*, 180:1–41, 2010.
- [17] Nicolai Haydn, Matthew Nicol, Sandro Vaienti, and Licheng Zhang. Central limit theorems for the shrinking target problem. *J. Stat. Phys.*, 153(5):864–887, 2013.
- [18] Hubert Hennion. Sur un théorème spectral et son application aux noyaux lipchitziens. *Proc. Amer. Math. Soc.*, 118(2):627–634, 1993.
- [19] I. A. Ibragimov and Yu. V. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen, 1971. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.
- [20] C. T. Ionescu Tulcea and G. Marinescu. Théorie ergodique pour des classes d’opérations non complètement continues. *Ann. of Math. (2)*, 52:140–147, 1950.
- [21] Yuri Kifer. Nonconventional Poisson limit theorems. *Israel J. Math.*, 195(1):373–392, 2013.
- [22] J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.*, 15(1):211–225, 1922.
- [23] J. Rousseau-Egele. Un théorème de la limite locale pour une classe de transformations dilatantes et monotones par morceaux. *Ann. Probab.*, 11(3):772–788, 1983.
- [24] Damien Thomine. A generalized central limit theorem in infinite ergodic theory. *Probab. Theory Related Fields*, 158(3-4):597–636, 2014.
- [25] Lai-Sang Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)*, 147(3):585–650, 1998.
- [26] Xuan Zhang. *Studies on the weak convergence of partial sums in Gibbs-Markov dynamical systems*. PhD thesis, The Pennsylvania State University, 2015.